Università degli Studi del Molise
Facoltà di Economia
Dipartimento di Scienze Economiche, Gestionali e Sociali
Via De Sanctis, I-86100 Campobasso (Italy)

ECONOMICS & STATISTICS DISCUSSION PAPER
No. 47/08

Family Income and Students' Mobility

by

Claudio Lupi
University of Molise, Dept. SEGeS

and

Patrizia Ordine
University of Calabria, Dept. of Economics and Statistics

Family Income and Students' Mobility*

Claudio Lupi†
University of Molise, Dept. of Economics, Management and Social Sciences
Patrizia Ordine‡
University of Calabria, Dept. of Economics and Statistics

August 27, 2008

Abstract

This paper investigates the reasons that determine students' mobility in Italy and tries to explain why in the presence of quality differentials among universities the majority of students choose to remain in their regions of origin. We find that low mobility is related to family income and other financial and background characteristics. Low mobility in turn implies the existence of little competition among universities, and hence little incentive for improvement in either teaching or research. A crucial issue is therefore to evaluate if and how the government may affect this process and improve the supply of higher education quality and the degree of competition among academic institutions.

JEL classification: I21, I28,
Key words: Higher education, University choice, Liquidity constraints.

1 Introduction

It is well understood in the economic literature that students' regional migration is relevant for designing correct educational policies. On this respect, educational policies need to set the appropriate level of resources to be devoted to higher education programs and to match qualification requirements needed by the economic activities in the region with the number and characteristics of the graduates. This topic has recently received special attention, since it has been recognized that the geographical mobility of students influences not only regional education policies but also the development and the growth path of local economies, as well as regional convergence. In particular, the theme of students' mobility

*An early version of this paper was presented at the conference "Education quality, institutions, and financial constraints", University of Turin, May 24th, 2006. We would like to thank, without implicating, the participants at the conference and Daniele Checchi for insightful comments.

†University of Molise, Dept. of Economics, Management and Social Sciences (SEGeS). Via F. De Sanctis. I-86100 Campobasso (Italy). lupi@unimol.it

‡Corresponding author. University of Calabria, Dept. of Economics and Statistics. Via P. Bucci. I-87036 Arcavacata di Rende, Cosenza (Italy). p.ordine@unical.it

has been extensively studied in the international literature aiming at evaluating to what extent individual, socio-economic and quality factors may determine the propensity to move (Baryla and Dotterweich, 2001; Dotterweich and Baryla, 2005; Hsing and Mixon, 1996; Sà et al., 2006).

At a microeconomic level, the individual choice of the university on the part of college students is driven by expected utility considerations and influenced by pecuniary and non-pecuniary costs. However, the specific factors influencing students' mobility are not clearly identified, but it is reasonable to think that the decision of moving is related to multi-level variables featuring the local labour market and the behavior of the economy, the family background and the households' financial endowment, the students' characteristics and the supply and quality of university facilities in the origin area and in the region of destination.

Taking the case of Italy, existing evidence (Brunello and Cappellari, 2005) suggests that the returns from college education are higher in the North of the country, but also that there is low students' inter-regional mobility. Hence, although employment probabilities and expected wages could be higher for students attending northern universities while monetary costs and fees are similar in the North and in the South of the country, there is low mobility among areas. Other studies also address the issue of students' mobility in Italy, finding that both university characteristics and socio-economic conditions of the geographical areas in which individuals are located are important (Agasisti and Dal Bianco, 2007).

This paper focuses on the internal regional mobility of Italian college graduates. We evaluate how students' inter-regional mobility is related to different factors in four Italian macro-areas, exploiting the properties of a varying-parameter model. Therefore, we explore which are the factors influencing the decisions of students to move in regions different from the region of origin. In particular, we evaluate the effect of family income and other financial and background characteristics on the students' choice to attend universities located in a region different from the region of residence.

The empirical analysis is developed in a Bayesian framework. The reason of this choice is threefold. First, it allows us to estimate a rather flexible varying-parameters model, where parameters are not constant across macro-areas. Second, by using a Bayesian framework we can carry out inference on quantities that combine both estimated parameters and data in a straightforward fashion. This feature allows us to comment our empirical findings directly in terms of the (probability) mean marginal effects of the explanatory variables, not in terms of the parameters values. In our opinion this is a big advantage. Last but not least, given that we can derive by simulation an approximation to the exact finite sample posterior densities of the parameters of interest, we don't need to rely on asymptotic properties in carrying out our inferences, nor we have to rely on unobserved data.[1]

The next two Sections are devoted to an illustration of the individual choice problem and its relations with the family and local background. The empirical analysis is carried out in Section 4. The last Section concludes.

[1]Of course, there are other theoretical reasons why one could prefer a Bayesian framework. The interested reader can refer to Appendix 1 in Lancaster (2004) for a quick summary.

2 Modelling Students' Mobility

Theoretical studies investigating the determinants of students' mobility focus on individual and socio-economic background as well as on geographical distance and education quality in an expected utility maximization framework (Mixon, 1992; Mixon and Hsing, 1994; Frenette, 2006). We set a theoretical reference model where movements between the origin regions o and the destination regions m are based on a comparison of net expected returns subject to budget constraints. If:

$$V_{io}^{m} = \int_{0}^{\infty} \left\{ \psi_{io}^{m} \int_{0}^{\overline{w}} w dF_{w}^{m} \right\} e^{-rt} dt - C_{io}^{m} \tag{1}$$

is the expected value of moving for the i individual from region o, $o = 1, 2, ...k$ to region m, $m = 1, 2, ...k$, r is the discount factor, ψ_{io}^{m} represents the probability of finding a job having moved for studying. In case of employment the expected real wage is given by $\int_{0}^{\overline{w}} w dF_{w}^{m}$ which depends on the wage distribution F_{w}^{m}. The cost of moving from the origin region can be written as a function of several components:

$$C_{io}^{m} = C\left(M_{io}, O_{io}, D_{io}\right) \tag{2}$$

where M_{io} are direct pecuniary costs, including university fees, O_{io} may be considered a fixed cost of leaving the origin region which depends on a number of individual characteristics such as age, marital status, home ownership and family socioeconomic background. D_{io} are additional variable costs which may depend on the region of destination and distance. For example, psychological costs possibly related to distance or information costs related to the existence of social networks.

The expected value of studying in the region of origin may be represented as:

$$V_{io}^{o} = \int_{0}^{\infty} \left\{ \psi_{io}^{o} \int_{0}^{\overline{w}} w dF_{w}^{o} \right\} e^{-rt} dt \tag{3}$$

where for simplicity we assume that $C_{io}^{o} = 0$. The individual moves when:

$$V_{io}^{m} - V_{io}^{o} > 0 \tag{4}$$

unless budget constraints are binding and the credit market is imperfect. In practice the following condition must also hold:

$$M_{io} \leq Y_{h} + A_{h} \tag{5}$$

where Y_h is the household's disposable income and A_h are available assets.

Indeed, there is no need for the budget constraint to be strictly binding for income and wealth to be important in formulating the choice of the university to be attended. Income and wealth can enter the decision formation process directly as components of the individual's utility function as

$$U_i = f\left\{(V_{io}^m - V_{io}^o), (Y_h + A_h - M_{io})\right\} \tag{6}$$

where utility of moving, U_i, is positively related to both components.

To sum up, this framework predicts that the decision of moving depends on individual, households and local characteristics. On top of that, we consider the possibility that parameters relating individual and household variables with the probability of moving are not constant across macro-areas. In practice. we explicitly consider how individuals with similar attributes, living in households with comparable characteristics, may have a different behavior in different macro-areas.

3 Mobility determinants

The basic theoretical set-up outlined in the previous Section states that the decision of attending a specific university, in the origin region or outside, depends on variables which shape the individual's net expected utility. This is determined by the future employment opportunities, by the expected monetary returns and by the pecuniary and non pecuniary costs. These variables in turn depend on the characteristics of students and households as well as on the socioeconomic macro environment.

3.1 Individual and household's characteristics

Individual and household characteristics affect individual mobility decisions since they determine the specific expected utility values and migration costs. In particular, individual-level covariates encompass gender, age, birth order, race/ethnicity, urban/rural origins, marital status, role in the family. The educational status and the individual ability, that can be measured by grades obtained in high school, may also determine the probability of moving for studying. Interestingly, even gender ma be considered a variable which can influence the decision of moving. In the recent migration literature (see e.g Pessar and Mahler, 2003; Hondagneu-Sotelo, 2003) it has been pointed out how a seemingly gender-neutral process of movement is, in fact, highly gender-specific and may result in differential outcomes for men and women. The interaction of women's roles, status, and age within a sociocultural context results in a specific propensity to migrate.

Family factors include size, age/sex composition, structure (nuclear, extended, etc.), status (single parent, both parents, etc.), and class status. The family educational background has always been recognized as a crucial variable in determining the investment in human capital and the schooling decisions (see e.g. Shea, 2000; Carneiro and Heckman, 2002; Checchi, 2003). Students from families with a higher cultural background tend to attend high quality schools/universities and to have higher propensity to move. As long as liquidity constraints may be binding in determining the decision to move, households' income and financial position (including home ownership) need to be considered when evaluating the choice of studying outside the origin region. The recent literature (Carneiro and Heckman, 2002) points out how the long-term effects of living in a wealthy family override the basic direct short-term impact since they have an effect on students' general abilities.

3.2 Macro determinants and local characteristics

Macro level covariates vary across regions and are supposed to capture observed utility differences between alternative destinations as suggested by the theoretical framework developed in Section 2. The characteristics of both the region of origin and the region of destination can influence the mobility propensities. Moreover, these characteristics can interact with the individual and household relations and affect decisions about studying location. Macro covariates include: the state of the economy (agrarian, industrial, the level of development); labour market conditions and conditions of work (wage levels, benefits); the ability of the economy to provide jobs and the type of jobs available; the number of industries; the ability of the national and local government to provide related infrastructure (education, job training); the geographic location of the destination region with respect to the native region. Social factors include those community norms and cultural values that determine whether or not student can migrate and, if they can, how (i.e., labour or family reunification) and with whom (alone or with the family).

4 The data and the statistical model

The data are taken from the Bank of Italy's survey of Italian households' income and wealth (SHIW). Indeed, SHIW is not specifically intended to be used to investigate education-related decisions. In this respect, other data sources may be more appropriate.[2] However, using the information contained in SHIW, we can merge education-related variables with financial information on individuals and households. To the best of our knowledge, this is a unique feature of SHIW. In order to have a larger number of observations, we use two recent waves of the survey (2002 and 2004: see Bank of Italy, 2004, 2006), being careful not to include the same individuals twice when they are present in both waves. Unfortunately, the most recent wave of the survey (which refers to 2006) does not include the relevant information for our enquiry.

Selected individuals are all graduated, so that we are carrying out a *conditional* analysis. Furthermore, they are not themselves head of households. Given the structure of the survey, this step was necessary in order to identify movers from non-movers. Of course this can be considered as a limitation to the analysis. However, the vast majority of Italian graduates fall in this category and we believe that we can gather important information even in the presence of data limitations.

[2] One such example is Istat (2005). However, this data source suffers from other problems. Apart from not reporting information about income and wealth, Istat (2005) does not allow to identify the region of residence of the movers. Indeed Istat (2005) tend to focus more on the decision of moving *after* graduating. Furthermore, regretfully some interesting answers to the questionnaire are blanked to non-Istat users.

Variable	Description
y	Dependent variable: 1 if the individual moved to a different region, 0 otherwise.
x_0	Intercept.
x_1	Low income: household's per-capita income less than the first quartile.
x_2	High income: household's per-capita income more than the third quartile.
x_3	Low-educated head of household (up to middle school).
x_4	High-educated head of household (university degree or higher).
x_5	Gender: 1 if female, 0 otherwise.
x_6	The household lives in a small city (less than 40,000 inhabitants).
x_7	The head of household is older than 64.
x_8	Number of universities in the region of residence.

Table 1 – Variables used in the analysis.

Prior to statistical analysis, the data are cleaned and missing observations are eliminated. Therefore, we are left with 523 individuals, about 18% of which decided to move to a different region[3] to study and obtain their university degrees.[4] Overall, the percentage of movers in Italy is very low as compared to other European countries. In fact, Naylor et al. (2001) report that about 88% of the college students in the UK chose to study in a different region with respect to where the parental home is located.

It is interesting to notice that, according to our data, the percentage of movers is highly variable across macro-areas, being 9.4% in the North-West, 15.6% in the North-East, 21.3% in the Centre, and 20.4% in the South. A detailed description of the variables used in the analysis is reported in Table 1[5] and some descriptive statistics are offered in Table 2.

Of course, the percentage of movers by macro-area is estimated as $100\, n_y/n$, with n_y the number of movers and n the sample size, from macro-area specific samples. Given that the sample size of macro-area specific subsamples is limited, one might wonder which is

[3]In this paper we denote by "region" the official NUTS2 classification. By "macro-area" we denote the official NUTS1 classification with the modification of considering "South" and "Islands" as a whole geographical entity that we call "South". However, the classification of the regions (NUTS2) in the four macro-areas (North-West, North-East, Centre and South) of the country follows the official classification used by the Italian National Institute of Statistics (ISTAT).

[4]The definition of "mover" is related to the decision of moving to a different region, independently of the region being in the same macro-area (North-West, North-East, Centre or South) or not. The aggregate figure we obtain from the Bank of Italy's surveys (17.6%) is broadly consistent with the analog ratio estimated using Istat (2005) data and with the aggregated proportion of movers (19.2%) computed on data made available by the Ministry of the University for the academic year 2003/2004.

[5]Other variables have been tentatively used in different parameterizations of the model that included households' net wealth and different proxies for students' quality and ability (including high school graduation marks and university degree final grades). However, they did not prove useful to explain individual choices, being always largely insignificant. Therefore, they have been excluded from the final model. Explanatory variables related to the destination region pose some problems of interpretability, in particular as far as the non-movers are concerned. To take into account the characteristics of both origins and destinations, a full flow analysis should be performed, which is beyond the scope of the present paper.

	NW	NE	CE	SO	Italy
Movers	9.4	15.6	21.3	20.4	17.6
Low income	13.5	15.6	22.1	48.5	29.1
High income	34.4	36.7	18.9	13.8	23.5
Low educated	54.2	42.2	53.3	42.9	47.2
High educated	16.7	12.8	12.3	26.0	18.4
Female	55.2	48.6	51.6	59.7	54.7
Small city	18.8	26.6	25.4	19.9	22.4
Elder head of household	15.6	12.8	16.4	21.4	17.4

Table 2 – Descriptive statistics (percentages).

	Mean	Low	High	$\Pr(\theta_{\text{NW}} < \theta_i)$
NW	10.1	5.3	15.0	
NE	16.3	10.6	21.9	90.6
CE	21.7	15.8	27.7	99.1
SO	20.5	15.9	25.1	99.2

Table 3 – Interval estimation of the percentage of movers in the population by macro-area. "Mean" represents the average of the posterior distribution of the parameter, "Low" and "High" are the lower and upper limit of the 90% highest posterior density interval (HPDI), respectively. $\Pr(\theta_{\text{NW}} < \theta_i)$ denotes the probability that the proportion in the North-West is less than the proportion in the specified macro-area. These values have been obtained using 10,000 simulations.

the percentage in the population (in the statistical sense) instead. Therefore we estimate the proportion of movers in the population by macro-area by using a simulation approach. In particular, denoting by θ the unknown population proportion of movers, we assume an uninformative prior such that $\theta \sim \text{U}(0, 1)$. With this prior we are simply assuming *a priori* that the true proportion is somewhere between 0 and 1 with equal probabilities. Under independence and constant θ it can be shown (see e.g. Gelman et al., 2004) that the posterior for θ is

$$\theta | n_y \sim \text{Beta}(n_y + 1, n - n_y + 1) \,. \tag{7}$$

It is easy to simulate this distribution using macro-area specific values for n_y and n and obtain in this way an interval estimation of the percentage of movers in each macro-area. In particular, for each macro-area we estimate the *Highest Posterior Density Intervals*, that is the Bayesian confidence intervals (or *credibility* intervals) such that θ_L and θ_H determine the shortest interval for which $\Pr(\theta_L \leq \theta \leq \theta_H) = \alpha$ $(0 < \alpha < 1)$. In our computations we consider $\alpha = 0.9$. We also compute the probability that the true value of the proportion of movers in the North-West is less than the proportion of movers in the other macro-areas (see Table 3).

Even allowing for substantial sample uncertainty, the probability that the population percentage of movers can be considered as equal among macro-areas is indeed very low. In fact, the population percentage of movers appears to be substantially lower in the northern regions than it is in the central and southern part of the country.

We suspect that, because of differences in economic environments and social conventions, not only the average percentages of movers are different across macro-areas, but also the individual decision process may be different. To the best of our knowledge, there is just some anecdotic evidence in this direction but, as far as we know, no serious statistical evidence has been proposed so far. For this reason we believe that it is useful to build a flexible logit model where parameters are allowed to vary across macro-areas. Specifically, we build a varying-parameter logit model where the decision to move or not to move on the part of individual i in macro-area r is represented by

$$
\begin{aligned}
y_{r,i} &\sim \mathrm{Bern}(\pi_{r,i}) && (8)\\
\mathrm{logit}(\pi_{r,i}) &= \beta_{0,r} + \beta_{1,r}x_{1,r,i} + \ldots + \beta_{k,r}x_{k,r,i} && (9)
\end{aligned}
$$

where $\mathrm{Bern}(\pi_{r,i})$ is the Bernoulli distribution with parameter $\pi_{r,i}$. Note that, consistently with our idea that individual decision may differ across macro-areas, we assume area-specific parameters. Indeed, the model can be considered as a stylized multilevel model without predictors (for the parameters). Therefore, we don't need to arbitrarily pick up one region as the *reference* one, as we would have to do by using a classical regression with indicator variables.[6] In our model, the varying intercepts account for the (fixed) characteristics of the macro-area of origin.

Given that the number of groups (macro-areas) is small (indeed, just 4) and that there are many varying parameters, the option of setting and estimating the model using a Bayesian approach can be advantageous. Also, as already highlighted in the introduction, the Bayesian approach allows us to derive the finite sample distributions of the parameters, as well as the finite sample distributions of quantities involving both parameters and data. This allows us in particular to derive interval estimations for the *mean marginal effects* of each explanatory variable in a rather natural way. Therefore, we complete the model with fairly standard priors on $\boldsymbol{\beta}_r := (\beta_1, \ldots, \beta_k)$. In particular we assume

$$
\begin{aligned}
\boldsymbol{\beta}_r &\sim \mathrm{N}_k\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right) && (10)\\
\mu_{\beta,j} &\sim \mathrm{N}\left(0, 1000\right) \quad j = 0, \ldots, k && (11)\\
\boldsymbol{\Sigma}_\beta &\sim \mathrm{InvWish}(k+3) && (12)
\end{aligned}
$$

where $\mathrm{N}_k\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right)$ indicates the multivariate normal distribution with (vector) mean $\boldsymbol{\mu}_\beta$ and covariance matrix $\boldsymbol{\Sigma}_\beta$, $\mathrm{N}(\mu, \sigma^2)$ is the normal with (scalar) mean μ and variance σ^2 and $\mathrm{InvWish}(\nu)$ denotes the inverse-Wishart distribution with ν degrees of freedom.[7] These priors, beside being rather conventional, do not impose strong constraints on the parameters. Indeed, though the proposed priors on the β's are proper, they are also extremely vague. In

[6]Following the suggestion of an Editor, we also estimate a standard logit model with interactions between macro-areas and income and education variables. The results are qualitatively similar to those derived using the Bayesian model, but the coefficients have to be interpreted in terms of deviations from the reference area (in our case, NW). Also, in a "classical" logit model it is more difficult to derive the finite sample distributions of the parameters and their probability implications. For all these reasons we prefer our Bayesian model. However, we report the results of the standard logit model in the appendix for completeness.

[7]The inverse-Wishart distribution is the conjugate prior distribution for the multivariate normal covariance matrix. For a $k \times k$ covariance matrix it is required $\nu > k$.

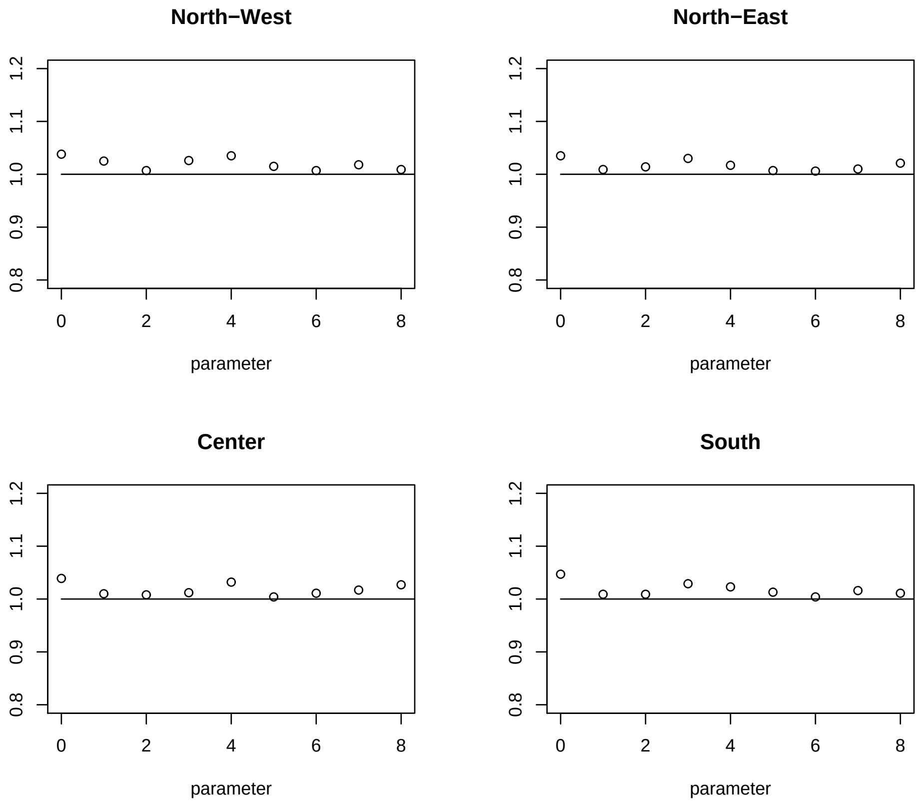

Figure 1 – Potential scale reduction factors (Gelman and Rubin, 1992). The x's indicate the parameter (0 for the constant and so on — see Table 1). Values close to 1 indicate convergence. All the results are obtained from 6 chains of effective length 1000.

order to be realistic, we use a non-diagonal covariance matrix $\boldsymbol{\Sigma}_\beta$ to take into account that the β's are not independent. However, as a robustness check, simpler independent priors on the β's have also been specified. In particular, we assumed the simple independent prior $\beta_{r,k} \sim \mathrm{N}(0, 1000)\ \forall r, k$. The results do not differ substantially from the proposed model.[8]

The model has been estimated by Monte Carlo Markov Chains (MCMC) using `R` and `WinBUGS` (R Development Core Team, 2006; Spiegelhalter et al., 2004). Six independent chains of length 1000 (excluding the burn-in replications) have been used to derive the posterior distributions of the parameters. The starting values of the chains were randomly selected from a $\mathrm{N}(\hat{\boldsymbol{b}}, \boldsymbol{I}_k)$ distribution, with $\hat{\boldsymbol{b}}$ representing the nation-wide estimates of a standard logit model.

Following Gelman et al. (2002), most of the results are reported graphically. However, for completeness and readability, we also report a table with the numerical values of the estimated parameters (Table 4).

Convergence was quickly reached for all the parameters of the model, as indicated by the potential scale reduction factors (Gelman and Rubin, 1992) reported in Figure 1.

Given that the random coefficient model is substantially more complicated than a stan-

[8]However, on the basis of the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), the original model has to be preferred. In fact, the DIC of the original model is 452.6, while the DIC of the model with independent priors is 468.5, leading one to reject the model with independent priors in favour of the complete model (see Spiegelhalter et al., 2002, p. 613).

dard logit model, in order to check if considering varying parameters is really useful, a standard logit model has also been estimated in a Bayesian framework and the Deviance Information Criteria (DIC, Spiegelhalter et al., 2002) of the two models have been compared. The DIC for the standard logit was 469.9, while we obtained a substantially lower value (452.6) for the random coefficients logit. According to Spiegelhalter et al. (2002, p. 613), this result should be interpreted as being strongly in favour of the model with varying parameters.

In order to check if the model fits the data well, a possibility is to perform a posterior predictive simulation analysis. In practice, one has to compare the observed data to replicated data sets simulated using the model (see e.g. Gelman et al., 2000; Geweke and McCausland, 2001). In particular it is possible to simulate for each individual

$$y_{r,i}^{repl} \sim \text{Bern}\left(\text{logit}^{-1}\left(\boldsymbol{x}_{r,i}\,\boldsymbol{\beta}_r^l\right)\right) \tag{13}$$

where $\boldsymbol{\beta}_r^l$ is a draw from the posterior distribution of $\boldsymbol{\beta}_r$ and $\boldsymbol{x}_{r,i}$ is the vector of explicative variables relative to individual i in region r. By simulating (13) for each individual 5,000 times, we obtain 5,000 replicated samples. If the model adequately describes the data, then the simulated data must mimic the observed ones.

An obvious check regards the fraction of movers for each macro-area. The results are reported in Figure 2 that represents the posterior predictive distributions of the proportion of movers in each macro-aea, as compared to the observed proportions. Notice that the model fits the data well in this respect: in fact, the average proportions computed over the simulated data are very close to the actual proportions of movers. Furthermore, the two-sided p-values are always very high, indicating that the proportions of movers implied by the model and those observed in reality are well in accordance.[9]

However, Figure 3 suggests that the model fits the data better for the central and southern regions than it does for the northern ones. In particular, in the North-west only a few observations where $y_i = 1$ are available and the range of fitted probabilities is small, indicating a likely poor predictive performance of the model in those regions. In this respect, it is fair to say that some of the best-renowned universities in Italy are located precisely in the north-western regions. Also, only a few potential outliers are apparent in those areas in which there is a relatively large number of observations. However, it should be noticed that, precisely because of the varying number of available observations, this is an expected result.

Estimated parameters are reported in Table 4 and graphically in Figure 4 using a common scale for ease of interpretation.

However, given that logistic regression coefficients do not have a direct probability interpretation, we compute also the *mean marginal effect* of each explanatory variable.

Making inferences on the mean marginal effects is usually a difficult task in a non-

[9] Denoting with $T(\boldsymbol{y}, \beta)$ the quantity of interest (in our case the proportion of movers) the two-sided p-value is defined as $2\min(p, 1-p)$, where $p := \Pr\left[T(\boldsymbol{y}^{repl}, \boldsymbol{\beta}) > T(\boldsymbol{y}, \beta)|\boldsymbol{y}, \boldsymbol{\beta}\right]$.

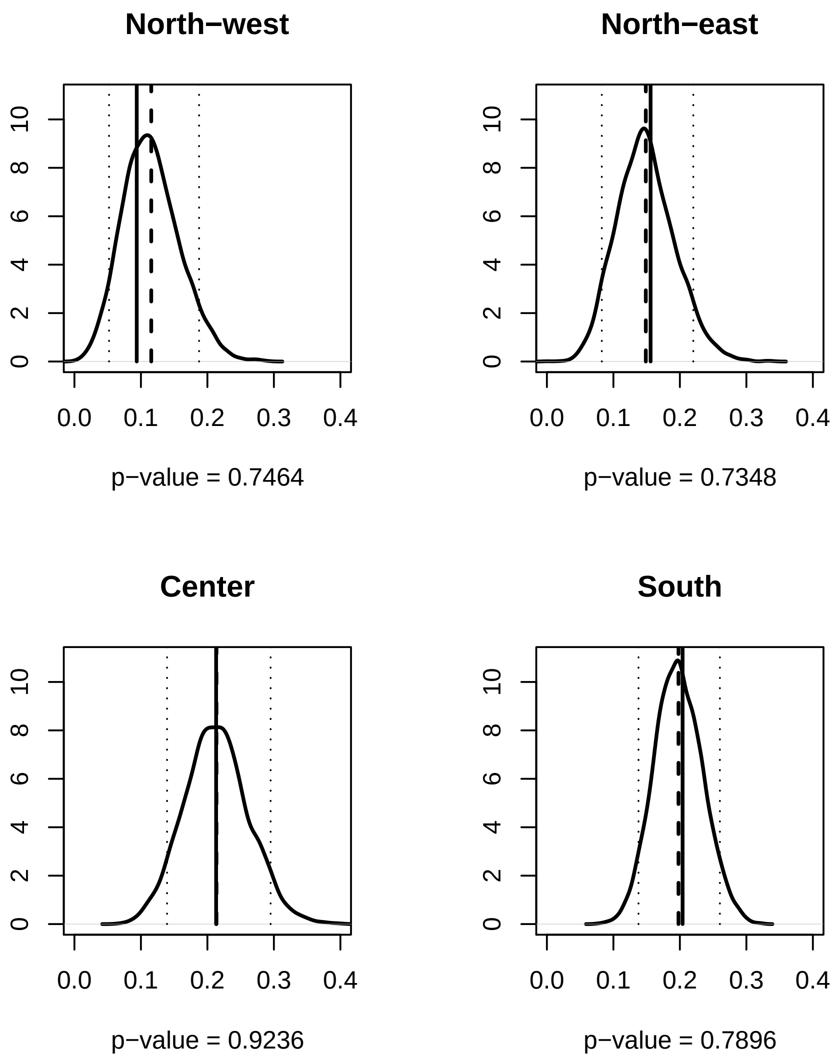

Figure 2 – Posterior predictive distributions of the proportion of movers by region. The solid vertical lines represent the observed proportions. The dashed vertical lines are the means of the posterior predictive distributions of the proportion of movers. The dotted lines correspond to the 5% and 95% quantiles of the distributions. "p-value" denotes the two-sided p-value.

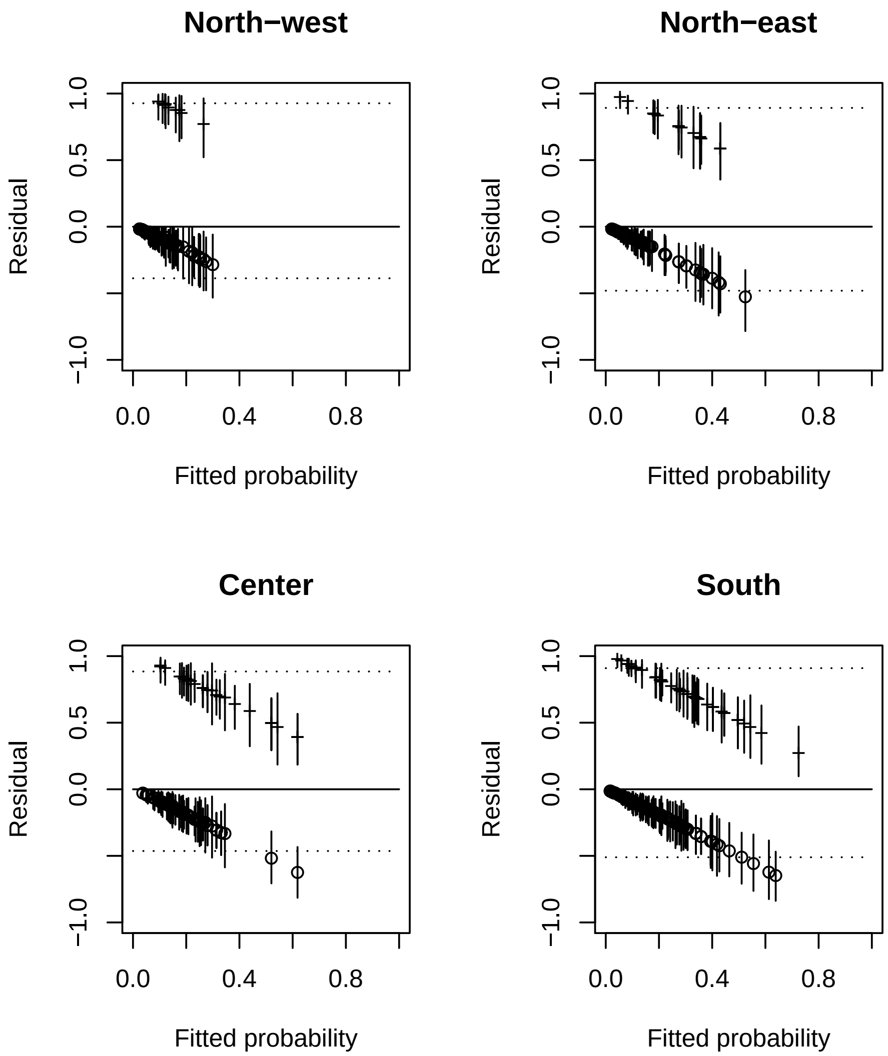

Figure 3 – Posterior distributions of the residuals $r_i = y_i - p_i$ plotted against $\mathrm{E}(p_i|y)$, with $p_i := \Pr(y_i = 1)$ (Albert and Chib, 1995). The vertical lines delimit 90% Highest Posterior Density Intervals (HPDIs). Circles denote medians of the observations where $y_i = 0$. Crosses indicate medians of the observations where $y_i = 1$. The dotted lines are the 2.5 and 97.5 percentiles of the distribution. Observations that significantly cross the dotted lines correspond to potential outliers.

Parameter	Area	Mean	Low	High
Intercept	NW	-1.4806	-2.6180	-0.2842
	NE	0.0015	-1.0230	0.9147
	CE	-0.4599	-1.4930	0.5194
	SO	-0.0359	-0.8069	0.7536
Low income	NW	-0.0053	-1.1130	1.2690
	NE	-0.9125	-1.8890	0.0085
	CE	-0.7816	-1.5450	0.0501
	SO	-0.9346	-1.5920	-0.2670
High income	NW	-1.0033	-2.0780	0.1041
	NE	-0.0287	-0.8064	0.7554
	CE	-0.3846	-1.1480	0.4235
	SO	-0.0492	-0.7306	0.6411
Low education	NW	-0.7286	-1.6990	0.2168
	NE	0.2527	-0.4945	0.9817
	CE	-0.2533	-0.8916	0.4272
	SO	0.1467	-0.4861	0.8275
High education	NW	0.5668	-0.4695	1.6390
	NE	0.9825	0.0740	1.8400
	CE	1.3422	0.4594	2.1250
	SO	0.8316	0.1722	1.5620
Gender	NW	0.4616	-0.4488	1.3570
	NE	0.4107	-0.2352	1.1220
	CE	0.4201	-0.1982	1.0280
	SO	0.1855	-0.3679	0.7555
Small city	NW	0.0575	-0.9387	1.1600
	NE	0.5765	-0.1901	1.3130
	CE	0.0776	-0.5911	0.7570
	SO	0.5054	-0.0831	1.1280
Age > 64	NW	-0.6144	-1.7510	0.5249
	NE	-0.4808	-1.4210	0.4026
	CE	-0.8945	-1.7360	-0.0229
	SO	-0.5400	-1.2340	0.1690
Number of universities	NW	-0.0546	-0.1566	0.0516
	NE	-0.6441	-0.9554	-0.3529
	CE	-0.1568	-0.2960	0.0060
	SO	-0.3743	-0.5391	-0.2059

Table 4 – Estimated parameters. “Mean” represents the average of the posterior distribution of the parameter, “Low” and “High” are the lower and upper limit of the 90% highest posterior density interval (HPDI), respectively.

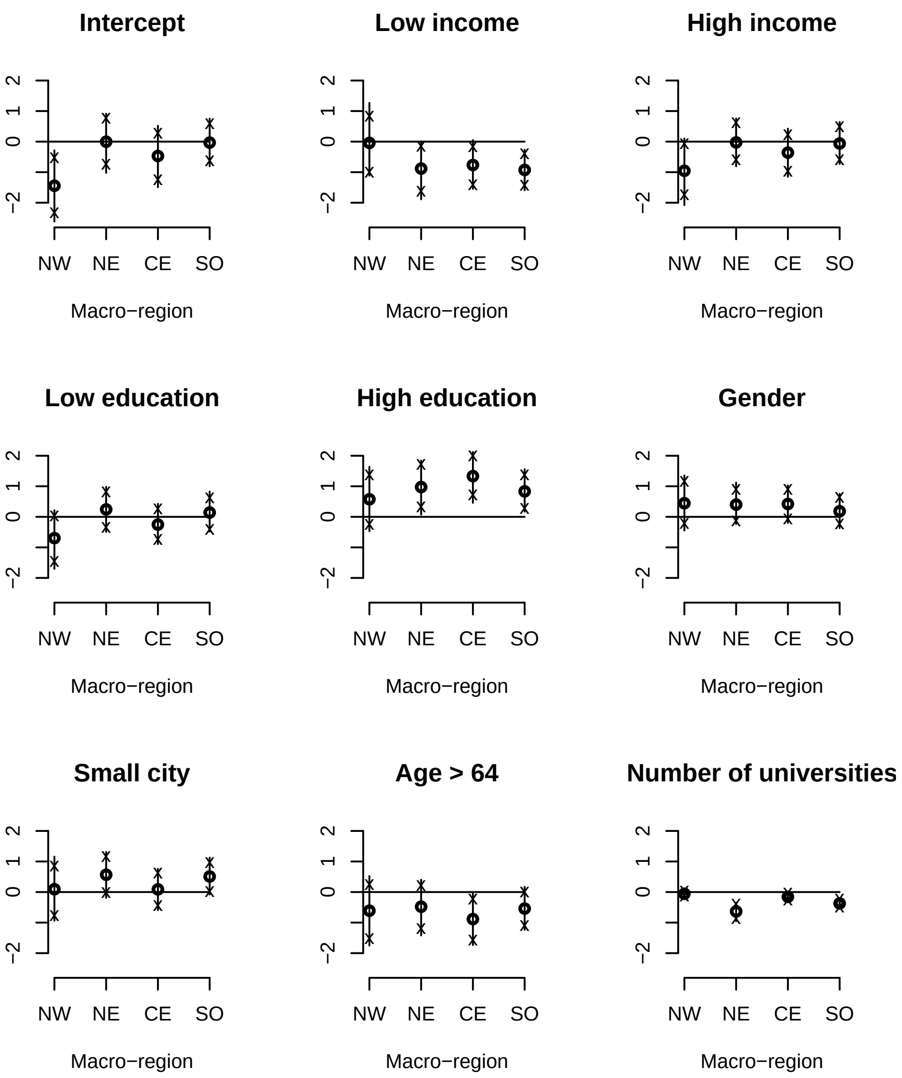

Figure 4 – Estimated parameters. The dots indicate the medians of the posterior distributions. The vertical lines and the "×" represent 90% and 80% highest posterior density intervals (HPDIs), respectively. Macro-areas are indicated using their abbreviations (NW, NE, CE, SO). The same scale is used for all parameters for ease of interpretation. Numerical values are listed in Table 4.

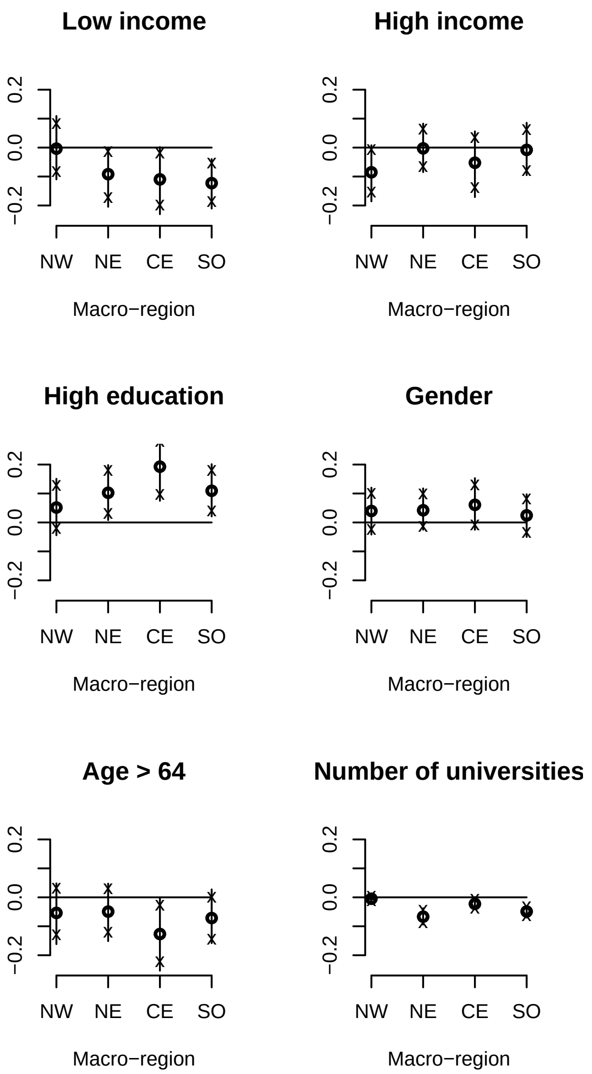

Figure 5 – Mean marginal effects. The dots indicate the medians of the posterior distributions. The vertical lines and the "×" represent 90% and 80% highest posterior density intervals (HPDIs), respectively. Macro-areas are indicated using their abbreviations (NW, NE, CE, SO). The same scale is used for all mean marginal effects for ease of interpretation.

Bayesian framework, since they involve both parameters and data. In particular, the mean marginal effect of the j-th explanatory variable over a sample of n individuals in a logit model with explanatory variables $\boldsymbol{z}_i := (z_{i1}, \ldots, z_{ik})'$ ($i = 1, \ldots, n$) and parameters $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_k)$ can be written as

$$n^{-1} \sum_{i=1}^{n} \frac{\partial \Pr(y_i = 1)}{\partial z_{ij}} = \gamma_j \, n^{-1} \sum_{i=1}^{n} \frac{e^{\boldsymbol{z}_i' \boldsymbol{\gamma}}}{\left(1 + e^{\boldsymbol{z}_i' \boldsymbol{\gamma}}\right)^2} . \tag{14}$$

Using a Bayesian approach, the posterior of this quantity can be computed directly from the posteriors of γ, thus allowing us to derive the distribution of each marginal effect and to carry out inference.[10]

The mean marginal effects are relatively uncertain for most variables, but some indications emerge quite clearly (see Figure 5).

In fact, the evidence suggests that family income exerts an important role in shaping individual decisions to move in most areas, but this influence seems related more to a possible budget constraint effect rather than to a generic income effect. The role of financial constraints on students' mobility is becoming increasingly relevant in the recent debate. For example, Frenette (2007), using data for Canada, finds that students from lower-income families saw the largest increase in university participation following the creation of a local degree-granting institution. This is consistent with the notion that distance poses a financial barrier. Indeed, in our model, while low income tends to be associated with lower probabilities of moving (the probability decreasing on average by approximately 10-15%), high income does not appear to symmetrically increase the probability of choosing universities far from home. In this respect, our results seem broadly consistent with the findings of Frenette (2007). However, it should be noted that the negative low-income effect is important in all areas, except the North-West, where some of the most highly reputed universities are located.

The education of the head of household is generally important in increasing the probability of choosing a university in a different region, especially so in the central regions where it can increase the probability of moving by approximately 20%. The effect is important also in the north-eastern and southern regions, where the probability of moving increases by about 10%. Apparently, better-educated parents are more keen to invest in their children's human capital. However, this result possibly indicate also a "search for quality" effect on the part of the more educated households. It is possible that more educated parents are also more informed (or more able to elaborate an independent judgement) about colleges quality. Instead, a low level of education of the head of the household is not associated with a symmetrically negative effect. We feel that this result is interesting and broadly consistent with the international literature (see e.g. Shea, 2000; Carneiro and Heckman, 2002; Checchi, 2003; Siegfried and Getz, 2006).

[10] In the frequentist approach it is of course possible to compute a point estimate of the mean marginal effects. However, deriving interval estimates and evaluating the statistical significance of the effects is a much more difficult task.

Contrary to our expectations, gender does not appear to be associated with any particularly clear pattern. We expected that traditional views that tend to maintain females closer to the family might be important in shaping the probability of moving, especially in the southern regions. Instead, we could not find any such discriminatory pattern. Of course this is a good result, in cultural terms.

As far as the dimension of the town of origin is concerned, it seems that a positive association exists between living in a small town in the North-East or the South with the decision of moving.

The presence of an elder head of household tends to be associated with a lower probability of moving, especially in central and southern regions, where family links are traditionally stronger.

Finally, the number of universities in the region[11] of origin is obviously important, being the main supply indicator. Larger numbers of universities in the region mean more opportunity to choose on the part of the students, and corresponding lower probabilities of moving to another region to complete the studies.

5 Concluding remarks

This paper investigates the reasons that determine students' mobility, explaining why even in the presence of quality differentials among universities the students may choose to remain in their regions of origin. We find that the relatively low mobility of Italian university students is related to family income and other financial and background characteristics. In particular, although employment probabilities and expected wages seem to be higher for students attending Northern universities, while monetary costs and fees are similar in the North and in the South of the country, there exists a low mobility among areas. This could imply the existence of little competition among universities, and hence little incentive for improvement in either teaching or research. A crucial issue is therefore to evaluate if and how the government may affect this process and improve the supply of higher education quality and the degree of competition among academic institutions.

The existing debate considers that the reform of the educational system must confront the issues of the degree of governmental ownership, the degree of governmental subsidy and the degree of competition. Recently, in Italy the traditional highly centralized state control has decreased and there is no longer the strong control over universities that was present until some years ago. A large amount of autonomy has been granted and the faculties have much stronger influence on the governing of individual universities. However, this does not seem to have increased the competition or reduced the quality differentials among universities.

A crucial issue is to evaluate if the government may affect the quality by changing the level of spending for higher education. In particular, it is relevant to evaluate the right instruments to implement this policy: tax exemptions, educational vouchers or direct subsi-

[11] Note that this is the number of universities that are present in the region of residence, not in the macro-area.

dies to the universities. According to our model, if credit constraints are really binding or if households are exposed at the risk of being liquidity constrained, the first two instruments might be more effective in improving the quality of education by breaking local monopolies and increasing competition among universities, especially in those areas where the financial markets are less developed or the accessibility to the credit market on the part of the household is seriously limited.[12] However, appropriate loan policies are possibly not sufficient. Our results seem to suggest that in order to have relevant effects, these policies should perhaps be coupled with better information policies. In this paper we suggest that more educated parents are not only more prone to invest in their children's education, but also are more likely to be more informed or able to elaborate independent judgements on colleges' quality, perhaps because they themselves have been university students some years before, to say the least. It would be important that the results of serious evaluation exercises would be easily made available to anyone interested in choosing the university to attend. In other words, public information on actual university quality is important. The degree of competition among universities may be undoubtedly improved by raising the individuals' information about universities' characteristics and students' performance in the labour market. Serious research and teaching quality evaluation programs attached to sufficiently strong incentives should in principle be able to increase universities competition both in terms of recruitment and capacity of attracting the best students. Of course, the results of these evaluation exercises should be made easily available and easily interpretable to the general public, so that the "search for quality effect" does not remain confined to the best informed households only. At present, there is only little experience in evaluating the Italian educational system and in publishing the evaluation's results.[13]

However, we believe that the monopoly power of local universities may be regulated effectively also by modifying structural characteristics and institutions operating in the regional economies. Inefficiencies in the credit market and liquidity constraints may induce high mobility costs among areas and consequently influence the supply of quality.

References

Agasisti, T., Dal Bianco, A., 2007. Determinants of college student migration in Italy: Empirical evidence from a gravity approach, paper presented at the Congress of the European Regional Science Association, Paris.

Albert, J. H., Chib, S., 1995. Bayesian residual analysis for binary response regression models. Biometrika 82 (4), 747–759.

Bank of Italy, 2004. Italian household budgets in 2002. Supplements to the Statistical Bulletin 12, Bank of Italy, Rome.

[12]The two aspect are closely related, but they are not identical. An indicator of financial development for the Italian regions has been estimated by Guiso et al. (2004). On the accessibility to the credit market, see Lupi (2005).

[13]The Italian university system is still largely based on rules rather than incentives. A lucid analysis is in Perotti (2003).

Bank of Italy, 2006. Household income and wealth in 2004. Supplements to the Statistical Bulletin 7, Bank of Italy, Rome.

Baryla, E., Dotterweich, D., 2001. Student migration: Do significant factors vary by region? Education Economics 3 (1), 269–280.

Brunello, G., Cappellari, L., 2005. The labour market effects of Alma Mater: Evidence from Italy. Discussion Paper 1562, IZA, Bonn.

Carneiro, P., Heckman, J. J., 2002. The evidence on credit constraints in post-secondary schooling. Economic Journal 112 (482), 705–734.

Checchi, D., 2003. The Italian educational system: Family background and social stratification. Working Paper 2003-01, University of Milan, Dept. of Economics.

Dotterweich, D., Baryla, E., 2005. Non-resident tuition and enrollment in higher education: Implications for tuition pricing. Education Economics 13 (4), 375–385.

Frenette, M., 2006. Distance to school and university participation. Education Economics 14 (1), 31–58.

Frenette, M., 2007. Do universities benefit local youth? Evidence from university and college participation, and graduate earnings following the creation of a new university. Studies Branch Research Paper Series 283, Statistics Canada.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis, Second Edition. Texts in Statistical Science. Chapman & Hall / CRC, Boca Raton, London, New York, Washington D.C.

Gelman, A., Goegebeur, Y., Tuerlinckx, F., Van Mechelen, I., 2000. Diagnostic checks for discrete-data regression models using posterior predictive simulations. Applied Statistics 49 (2), 247–268.

Gelman, A., Pasarica, C., Dodhia, R., 2002. Let's practice what we preach: Turning tables into graphs. American Statistician 56 (2), 121–130.

Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4), 457–472.

Geweke, J., McCausland, W., 2001. Bayesian specification analysis in econometrics. American Journal of Agricultural Economics 83 (5), 1181–1186.

Guiso, L., Sapienza, P., Zingales, L., 2004. Does local financial development matter? Quarterly Journal of Economics 119 (3), 929–969.

Hondagneu-Sotelo, P. (Ed.), 2003. Gender and U.S. Immigration: Contemporary Trends. University of California Press, Berkeley, CA.

Hsing, Y., Mixon, F. J., 1996. A regional study of net migration rates of college students. Review of Regional Studies 26 (2), 197–209.

Istat, 2005. Inserimento professionale dei laureati – indagine 2004. Manuale utente, Istituto Nazionale di Statistica, Rome.

Lancaster, T., 2004. An Introduction to Modern Bayesian Econometrics. Blackwell Publishing, Oxford.

Lupi, C., 2005. Are credit constraints in Italy really more binding in the South? Economics Bulletin 3 (35), 1–6.

Mixon, F. J., 1992. Factors affecting college student migration across States. International Journal of Manpower 13 (1), 25–32.

Mixon, F. J., Hsing, Y., 1994. College student migration and human capital theory: A research note. Education Economics 2 (1), 65–73.

Naylor, R., Boero, G., McKnight, A., Smith, J., 2001. Graduates and graduate labour markets in the UK and Italy. Lavoro e relazioni industriali 7 (2), 131–172.

Perotti, R., 2003. The Italian university system: Rules vs incentives, paper presented at the conference "Monitoring Italy", ISAE, Rome.

Pessar, P. R., Mahler, S. J., 2003. Transnational migration: Bringing gender in. International Migration Review 37 (3), 812–846.

R Development Core Team, 2006. R 2.4.0 - a language and environment for statistical computing. Tech. rep., R Foundation for Statistical Computing, Vienna.

Sà, C., Florax, R. J. G. M., Rietveld, P., 2006. Does accessibility to higher education matter? Choice behaviour of high school graduates in the Netherlands. Spatial Economic Analysis 1 (2), 155–174.

Shea, J., 2000. Does parents' money matter? Journal of Public Economics 77 (2), 155–184.

Siegfried, J., Getz, M., 2006. Where do the children of professors attend college? Economics of Education Review 25 (2), 201–210.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64 (4), 583–639.

Spiegelhalter, D. J., Thomas, A., Best, N., Lunn, D., 2004. WinBUGS user manual. Tech. rep., MRC Biostatistics Unit, Institute of Public Health.

Appendix: the standard logit with interactions

In this appendix we report the results deriving from the estimation of a standard logit model, where the coefficients related to income and education variables are allowed to interact with the macro-area indicator.

Though the coefficients of this model are not directly comparable with those of our Bayesian model for the reasons explained in the main text, nevertheless, the results are broadly consistent.

	Estimate	Std. Error	z value	Pr(> $\|z\|$)
Intercept	-0.4785	0.7742	-0.6180	0.5365
NE	-1.2255	0.8861	-1.3830	0.1667
CE	-0.2277	0.7838	-0.2910	0.7714
SO	-0.1101	0.7615	-0.1450	0.8850
Low income	1.6636	0.9396	1.7710	0.0766
High income	-1.4397	1.1585	-1.2430	0.2140
Low education	-1.8193	0.9597	-1.8960	0.0580
High education	-0.0134	0.9835	-0.0140	0.9891
Gender	0.3171	0.2563	1.2370	0.2160
Small city	0.4151	0.2834	1.4650	0.1430
Age > 64	-0.6586	0.3687	-1.7860	0.0740
Number of universities	-0.1581	0.0527	-2.9970	0.0027
Low income * NE	-2.8544	1.4549	-1.9620	0.0498
Low income * CE	-2.6993	1.1830	-2.2820	0.0225
Low income * SO	-2.7949	1.0546	-2.6500	0.0080
High income * NE	1.4516	1.3042	1.1130	0.2657
High income * CE	1.0076	1.3377	0.7530	0.4513
High income * SO	1.5754	1.2643	1.2460	0.2127
Low education * NE	2.5477	1.1462	2.2230	0.0262
Low education * CE	1.8295	1.1040	1.6570	0.0975
Low education * SO	1.9523	1.0820	1.8040	0.0712
High education * NE	1.2835	1.2661	1.0140	0.3107
High education * CE	2.3121	1.2099	1.9110	0.0560
High education * SO	0.3701	1.0900	0.3400	0.7342

Table 5 – Standard logit model with interactions.